Structure
Opinions

Generation of Protein Structures for the 21st Century

Raymond C. Stevens[1,*]
[1]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA
*Correspondence: stevens@scripps.edu
DOI 10.1016/j.str.2007.11.003

By their very nature, science and technology change over time. The answers to one question often lead to a host of new questions, and as technology improves we change not only how well we can address specific problems but even the types of problems we can address. This is particularly true for biomedical research over the past 50 years. Recent advances have allowed us to frame questions about evolution and development, and various types of molecular networks within cells that were previously inaccessible with more limited data collection and reductionist approaches. The change to large-scale data acquisition for DNA sequencing projects has been a productive one, and while one needs to be careful about changing for the sake of change, in general change is good. Based on the success of the genome projects, which also met with considerable resistance in the initial stages, a careful analysis was conducted over 10 years ago by NIGMS and NIH-funded researchers that clearly highlighted a need for more structural biology information than was currently being generated. Furthermore, the need for more structural data was only going to increase with time. The generation of one de novo structure every 2–3 years by each structural biology research group at an average cost in excess of 350K per new structure was simply not sufficient or efficient given the projected need.

High-Throughput Structural Biology and Technology Development

Lost in the Protein Structure Initiative (PSI) debate are the advanced structural-biology-specific technologies that have emerged. PSI-1 was primarily about developing, evaluating, and nurturing the technologies to ascertain if the rate of protein structure determination and the efficiency of the process could be increased and conducted in a production-like format similar to the genome projects. Thus, when one calculates the overall cost of PSI, a significant percentage was devoted to the technologies that were delivered which were of general use to all scientists, fulfilling important requirements that had not been previously addressed. In PSI-2, there are six centers focused on technology development, particularly for the most challenging proteins (e.g., membrane proteins, large and transient complexes, and eukaryotic proteins), and four centers focused on production, which have technology development components. Several of the PSI-1 and PSI-2 technology highlights that emerged either directly or indirectly include (complete list is available at http://cci.lbl.gov/kb-tech/):

- Improved cloning and expression tools specifically designed for structural biology;
- Increased biophysical understanding of protein constructs likely to crystallize, constructs better suited for NMR analysis, and which should not be pursued in their current state;
- Reduction in sample volume/concentration requirements, particularly for hard to express proteins (e.g., membrane proteins/complexes):
- Nanovolume crystallization, and
- Microcoil NMR;
- Crystallization and imaging robotics;
- Automated 1D NMR screening of samples (7 μl/20 min per sample);
- Synchrotron automated crystal sample changers;
- Improved NMR/X-ray data processing/structure solution software;
- Data management systems to evaluate both negative and positive results that improve our future experimental design;
- Improved metrics for structure quality.

Of critical importance, the robustness of the technologies have made the determination of protein structure more achievable by biochemists and cell biologists and the cost of the technologies have decreased to the point that they are now becoming commonplace. With miniaturization, reagent cost also goes down (similar to the change to microscale organic synthesis decades ago). It is quite likely that as more and more chemists and biologists conduct routine molecular biology in their own laboratories, a similar situation will occur with many protein structure projects being determined by "non-structural biologists". This will probably bring shivers to the traditional structural biologists who have enjoyed publishing just structures in high-profile journals, but it is a positive direction for science overall. A significant and different perspective is provided by what an enzymologist can see in the electron density of the active site or what a neurobiologist can see in the electron density of a channel structure.

More Effective Training of Students and Postdocs

Also lost in much of the debate concerning PSI is the impact on postgraduate education. One of the major drivers for many of those involved in PSI-1 was a frustration with the pace of obtaining structural data, hit-or-miss successes, and the repetitive nature of the existing technologies. In evaluation of my lab and other structural biology labs in 1995, I estimated that more than 60%

of the students time on de novo structure determination projects was spent conducting routine tasks repeatedly, such as crystallization trials with different constructs or purification procedures, looking for crystals to mount in the cold room, and then staying up for 24 hr straight at a synchrotron beamline walking in one sample at a time. In the case of de novo structure determinations, this situation did not allow for the majority of their time thinking about or understanding the function and mechanism of the beautiful structures they uncovered. If today, students and postdocs have to conduct structural biology in strictly the "old-fashioned" way, then this is indeed unfortunate for them. The argument will be made that having students and postdocs conduct research in production centers is a robotic-like task in itself, but the centers are very sensitive to this. Technicians conduct most of the routine tasks, while students and postdocs are encouraged to analyze and think about the information being generated.

Impact

As with the Human Genome Project, it is difficult to immediately appreciate the impact of PSI on biomedical research. As noted above, the technologies are key deliverables already in place from PSI. The integration of these technologies throughout the country and the world is the strongest possible endorsement of their critical need. While PSI was getting started and in the same spirit of PSI with analogous approaches to drug discovery, several high-throughput structure-based drug discovery companies were formed. These companies are all now maturing and are starting to yield results from their new "PSI-like" approaches to drug discovery. Specifically, Syrrx (now called Takeda San Diego) has a Phase III drug for type 2 diabetes in human clinical trials. SGX Pharmaceuticals and Astex Therapeutics both have drugs currently in Phase I and II human clinical trials for oncology. Likewise, pharmaceutical companies now value structural data so much that it is considered critical data prior to entering clinical trials, an important change in the last 5 years due in part to the PSI related technologies.

An example of impact in the academic setting is the recent structure of a human G protein coupled receptor (GPCR; Cherezov et al., 2007; Rosenbaum et al., 2007). GPCRs are the largest single family of proteins in the human genome and the target of more than 50% of all current marketed drugs. The structure determination is the result of the tireless pursuit of GPCR structure by biochemist Brian Kobilka at Stanford University coupled to PSI technology developed in my laboratory specifically for this particular family of receptors, namely miniaturization and automation of novel crystallization/imaging methods for GPCRs. This collaborative effort resulted in the long-awaited, high-resolution structure of human β_2 adrenergic receptor. Obtaining the three-dimensional structure of GPCRs has been the pursuit of a large number of groups around the world; both academia and industry have invested millions and millions of dollars into this single goal. Although one can argue that the above technologies and structures would have eventually emerged, PSI in partnership with the NIH Roadmap Initiative have accelerated all of this. Surely most biologists would like structural data on their newly discovered genes sooner rather than later.

Protein Initiative-3 (or 1)

I am personally not excited about solving structures with novel folds so I am not a participant in a PSI-2 production center. My interests are in structural neurobiology and in developing critically needed technologies that help myself and others accomplish their research goals. However, I do appreciate the passion the PSI-2 production centers have for their work and understand and appreciate the big picture that will result from their dedicated effort, as I also appreciate the enormous efforts of those working on viruses, ribosomes, and other very challenging structural biology projects. Once PSI-2 is completed in the summer of 2010, the issue of how we will generate the structural data that the scientific community needs will not go away. Protein structures will continue to be critical for the understanding of chemistry and biology. Key considerations for a Protein Initiative-3 (or 1) include:

- Careful re-evaluation of the targets/approaches for biomedical research.
- Continuation of technology development. This is critical for studying the larger complexes and challenging proteins like those that exist in the membrane. We also need increased production of these very challenging protein structures so that we improve our fundamental understanding of these key biological macromolecules.
- Function is as important as the structures being generated and this should become a more central focus of any future program. Already, movement is evident in this area with chemical profiling of protein families in a structural genomics setting (Allali-Hassani et al., 2007).
- Better coordination with other proteomics approaches (e.g., activity-based profiling, metabolite profiling of enzyme families, and mass spectrometry differentiation among various cellular species).
- Although we have seen X-ray and NMR complement one another better through PSI, electron microscopy has not been integrated as well, in part due to the proteins being studied. As we start to investigate larger complexes, the membrane, and entire cells, electron microscopy and single molecule analysis advances will be extremely important techniques to integrate and further our understanding of biological structures and their functions.

There is no question that certain things could be done better in PSI, as is often the case when change occurs. What is important is that we have made tremendous strides in improving our approaches to collect and understand protein structures in the past several years, and we are now in a better position to keep up with the demand for more structural data. I applaud NIGMS staff for talking to scientific leaders about current limitations and potential new directions in the field, and for

taking the initiative to change the way we do science based on prior investments and new discoveries.

REFERENCES

Allali-Hassani, A., Pan, P.W., Dombrovski, L., Najmanovich, R., Tempel, W., Dong, A., Loppnau, P., Martin, F., Thornton, J., Edwards, A.M., et al. (2007). Structural and chemical profiling of the human cytosolic sulfotransferases. PLoS Biol. *5*, e97.

Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., et al. (2007). High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. Science *318*, 1258–1265.

Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Yao, X.J., Weis, W.I., Stevens, R.C., et al. (2007). GPCR engineering yields high-resolution structural insights into β2 adrenergic receptor function. Science *318*, 1266–1273.

Update on the Protein Structure Initiative

John C. Norvell[1,*] and Jeremy M. Berg[1,*]
[1]NIGMS, NIH, 45 Center Drive MSC 6200, Bethesda, MD 20892-6200, USA
*Correspondence: norvellj@nigms.nih.gov (J.C.N.), bergj@mail.nih.gov (J.M.B.)
DOI 10.1016/j.str.2007.11.004

Based on the success of genome sequencing projects and advances in structural and computational biology, in the late 1990s scientists from several countries proposed large-scale projects to map protein structure space. The new field of structural genomics was defined as the high-throughput experimental determination of a large number of representative structures, with the goal of achieving systematic sampling of sequence families. Utilization of computational modeling of sequence family homologs would extend the structural information to a much larger fraction of sequenced genes. One of the national efforts, the Protein Structure Initiative (PSI), was established in 2000 by the National Institute of General Medical Sciences (NIGMS), NIH, as one of the special initiatives established during the NIH budget doubling and after several national and international workshops and extensive consideration by the Institute staff and the Institute's Advisory Council. The NIGMS held three workshops to examine the feasibility, goals, scale, and target selection strategy for a structural genomics effort. Following these workshops and staff discussions, the Council concluded that the Institute should undertake this effort and asked the NIGMS staff to organize a "pilot" phase of the PSI as a 5-year project with the mission statement: "To make the three-dimensional atomic level structures of most proteins easily available from knowledge of their corresponding DNA sequences."

The First Phase of the Protein Structure Initiative (PSI-1)

PSI-1 consisted of a centers program and an investigator-initiated grants program for methodology and technology development. Nine pilot research centers were established to test strategies for high-throughput structural determination. Two of these pilot centers were cofunded by the NIH National Institute of Allergy and Infectious Diseases (NIAID). The goals of PSI-1 were to:

1. Develop methodology and technology to increase success rates and lower costs of structure determination,
2. Construct and automate the protein production and structure determination pipeline, and
3. Determine novel protein structures. In this context, the term "novel" was defined to mean structures for proteins that were less than 30% identical in sequence to proteins for which structures had already been determined.

During the first year, the Institute appointed the Protein Structure Initiative Advisory Committee (PSIAC), a working group of the NIGMS Council composed of independent scientists (i.e., not connected to the PSI), to provide strategic advice to the NIGMS Council and staff on the management and planning of the project. In formulating this program, the intent was explicitly not to compete with traditional high resolution structural biology, but rather to generate a large body of novel structural information for use by the broad biomedical research community.

Over the five years of PSI-1, the nine pilot centers determined about 1300 structures of which approximately 65% were novel (based on the 30% sequence identity criterion). Structures contributed by PSI are comparable in quality and size to structures deposited into the Protein Data Bank (PDB) from other structural biology laboratories. Since these centers took several years to reach high-throughput operation, it was not surprising that 40% of the PSI-1 structures were determined in the fifth year of the project. By this time, the cost per structure had fallen more than 2-fold—to $138,000. This estimated cost per structure includes funds for ongoing technology development.

From this first phase of the PSI, NIGMS staff and the PSIAC concluded that several lessons had been learned:

- Structural genomics pipelines can be constructed and scaled up;